Table S1 GenBank accession numbers and references of organisms and proteins presented in main text and figures

Figure	Organisms Name	**Organism Accession Number**	Reference(s)
Fig 2	*Pseudomonas putida*	AB238971	Keil *et al*. 1985; Keil *et al*. 1987a; Keil *et al*. 1987b; Osborne *et al.* 1988; Assinder *et al.* 1992; Assinder *et al.* 1993; Gallegos *et al*. 1997; Sentchilo *et al.* 2000; Tsuda and Genka 2001; Yano *et al.* 2007; Miyakoshi *et al.* 2012
Fig 2	*Treponema primitia* str. ZAS-2	CP001843	Rosenthal *et al.* 2011
Fig 2	*Treponema primitia* str. ZAS-1	CP001843	Ballor *et al.* 2011
Fig 2	*Novosphingobium aromaticivorans*	AF079317	Romine *et al*. 1999
Fig 2	*Novosphingobium* sp.	FR856862	D'Argenio *et al*. 2011
Fig 2	*Methylocella silvestris*	CP001280	Chen *et al*. 2010
Fig 2	*Sphingobium japonicum*	AP010803	Nagata *et al.* 2010
Fig 2	*Azoarcus* sp.	AM406670	Krause *et al.* 2006
Fig 2	*Thauera* sp.	CP001281	NA
Fig 2	*Dechloromonas aromatica*	CP000089	NA
Fig 2	*Azotobacter vinelandii*	CP001157	Setubal *et al.* 2009
Fig 2	*Methylibium petroleiphilum*	CP000555	Kane *et al*. 2007
		Catechol 2,3-dioxygenase Accession Number	
Fig 3	*Novosphingobium aromaticivorans*	NP_049202	Romine *et al*.1999
Fig 3	*Sphingomonas agrestis*	AAB03075	Yrjala *et al.*1994
Fig 3	*Rhizobium* sp.	ABF82226	NA
Fig 3	*Beijerinckia* sp.	B57264	Kim and Zylstra 1995
Fig 3	*Sphingomonas* sp.	AAM14600	NA
Fig 3	*Sphingomonas* sp.	ADK27485	NA
Fig 3	*Novosphingobium pentaromativorans*	ZP_09195387	NA
Fig 3	*Novosphingobium* sp.	YP_004534265	D'Argenio *et al*. 2011
Fig 3	*Sphingomonas* sp.	AAD11448	NA
Fig 3	*Sphingomonas* sp.	AAD11452	NA
Fig 3	*Treponema primitia* str. ZAS-1	ZP_09718346	Ballor *et al.* 2011
Fig 3	*Treponema primitia* str. ZAS-2	YP_004531633	Rosenthal *et al.* 2011
Fig 3	lake sediment		
Fig 3	activated sludge	BAH89314	Suenaga *et al*. 2007; Suenaga *et al*. 2009
Fig 3	activated sludge	BAH90343	Suenaga *et al*. 2007; Suenaga *et al*. 2009

Fig 3	*Methylocella silvestris*	YP_002361794	Chen *et al*. 2010
Fig 3	*Sphingobium japonicum*	YP_003544642	Nagata *et al.* 2010
Fig 3	*Pseudomonas putida*	BAB62050	NA
Fig 3	*Thauera* sp.	YP_002890083	NA
Fig 3	*Dechloromonas aromatica*	YP_287003	NA
Fig 3	*Azotobacter vinelandii*	YP_002800217	Setubal *et al.* 2009
Fig 3	*Pseudomonas putida*	YP_709322	Keil *et al*. 1985; Keil *et al*. 1987a; Keil *et al*. 1987b; Osborne *et al.* 1988; Assinder *et al.* 1992; Assinder *et al.* 1993; Gallegos *et al*. 1997; Sentchilo *et al.* 2000; Tsuda and Genka 2001; Yano *et al.* 2007; Miyakoshi *et al.* 2012
Fig 3	*Alcaligenes xylosoxydans*		
Fig 3	*Marinobacter algicola*	ZP_01893284	NA
Fig 3	whale fall rib bone		
Fig 3	*Marinobacterium stanieri*	ZP_09507225	NA
Fig 3	*Marinobacterium stanieri*	ZP_09506167	NA
Fig 3	marine gamma proteobacterium HTCC2207	ZP_01224201	NA
Fig 3	sludge		
Fig 3	*Novosphingobium nitrogenifigens*	ZP_08210144	NA
Fig 3	*Cupriavidus necator*		
Fig 3	*Methyloversatilis universalis*	ZP_08506618	NA
Fig 3	*Methylibium petroleiphilum*	YP_001021468	Kane *et al*. 2007

Table S2 *Treponema primitia* str. ZAS-1 and ZAS-2 *meta*-cleavage pathway genes and pfam features

Gene Annotation	Present in str. ZAS-1	Present in str. ZAS-2	Protein Size (a.a.) in str. ZAS-1	Protein Size (a.a.) in str. ZAS-2	Pfam(s) Represented (PF#, family, domain)
ferredoxin-like peptide	+	+	98	94	PF00111, Fer2, [2Fe-2S] cluster binding
catechol 2,3-dioxygenase	+	+	308	308	PF00903, Glyoxalase, glyoxalase/ bleomycin resistance protein/ dioxygenase PF00903, Glyoxalase, glyoxalase/ bleomycin resistance protein/ dioxygenase
2-hydroxymuconic semialdehyde hydrolase	+	+	274	274	PF00561, Abhydrolase 1, a/b hydrolase fold
2-oxopent-4-enoate hydratase	+	+	260	260	PF01557, FAA hydrolase, fumarylacetoacetate hydrolase
4-hydroxy-2-oxopentanoate aldolase	+	+	340	335	PF00682, HMGL-like, HMGL-like PF07836, DmpG comm., DmpG-like communication
acetaldehyde dehydrogenase	+	+	291	288	PF01118, Semialdehyde dh, semialdehyde dehydrogenase NAD binding PF09290, Acetdehyd dimer, prokaryotic acetaldehyde dehydrogenase dimerisation
2-hydroxymuconic semialdehyde dehydrogenase	-	-			PF00171 Aldedh, aldehyde dehydrogenase family
4-oxalocrotonate tautomerase	-	-			PF01361 Tautomerase, tautomerase enzyme
4-oxalocrotonate decarboxylase	-	-			PF01557 FAA hydrolase, fumarylacetoacetate hydrolase
4-oxalocrotonate carboxy-lyase	-	-			PF01557 FAA hydrolase, fumarylacetoacetate hydrolase

Table S3 Gene expression patterns of *Treponema primitia* str. ZAS-2 in pure culture or co-culture with *T. azotonutricium* str. ZAS-9

Cohort	Expression Range	Pure Culture		Co-Culture	
	(reads/kb)	No. genes	%	No. genes	%
1[a]	0	344	9	395	10
2	1-9	503	13	556	14
3[b]	10-99	1642	43	1548	40
4[c]	100-999	1248	32	1234	32
5	> 1000	110	3	117	3

Analysis complements Table S4.
Analysis based on expression dataset from Rosenthal *et al*. 2011.
[a]In addition to having 0 reads/kb, all genes in Cohort 1 have 0 expression.
[b]Cohort includes *meta*-cleavage pathway 2-hydroxymuconic semialdehyde hydrolase, 2-oxopent-4-enoate hydratase, 4-hydroxy-2-oxopentanoate aldolase, acetaldehyde dehydrogeanse, and the associated ferredoxin-like peptide.
[c]Cohort includes *meta*-cleavage pathway catechol 2,3-dioxygenase.

Table S4 Expression of *Treponema primitia* str. ZAS-2 genes

Gene	Pure Culture Expression (normalized reads/kb)	Co-Culture Expression (normalized reads/kb)
<u>*Meta*-cleavage pathway</u>		
Ferredoxin-like peptide[a]	53	13
catechol 2,3-dioxygenase[b]	219	170
2-hydoxymuconic semialdehyde hydrolase[a]	43	60
2-oxopent-4-enoate hydratase[a]	60	43
4-hydroxy-2-oxopentanoate aldolase[a]	61	36
acetaldehyde dehydrogenase[a]	59	62
<u>House-keeping</u>		
ATP-dependent Clp protease ATP-binding subunit ClpX	314	345
Metallo-beta-lactamase family protein, RNA-specific	140	144
RNA polymerase sigma factor RpoD	72	77
RNA polymerase sigma factor RpoD	183	135
RNA polymerase sigma factor RpoD	188	162
RNA polymerase sigma factor RpoD	94	92
RNA polymerase sigma factor RpoD	30	94
<u>Acetogenesis</u>		
Formate dehydrogenase chain D	62	2
Formate dehydrogenase chain D	399	21
Formate dehydrogenase; cysteine-containing variant	601	33
Formate dehydrogenase; selenocysteine-containing variant	168	331
Carbon monoxide dehydrogenase CooS subunit	2698	2524
Formate-tetrahydrofolate ligase (FTHFS)	1463	1600
<u>Glycolysis</u>		
Hexokinase	168	190
Glucose-6-phosphate isomerase	169	170
6-phosphofructokinase	121	139
Fructose-bisphosphate aldolase class II	703	947
Triosephosphate isomerase	344	418
NAD-dependent glyceraldehyde-3-phosphate dehydrogenase	134	173
NAD-dependent glyceraldehyde-3-phosphate dehydrogenase	152	167
2,3-bisphosphoglycerate-independent phosphoglycerate mutase	440	393
Enolase	851	967
Pyruvate kinase	89	93
<u>Misc.</u>		
acetate kinase	914	1085
Fe-S cluster containing hydrogenase components 2	552	39
Periplasmic [Fe] hydrogenas large subunit	495	31

Expression data from Rosenthal *et al*. 2011.

Normalized gene expression values are number of reads mapped to a particular gene divided by size of that gene.
[a]Belong to Cohort 3 (Table S3).
[b]Belongs to Cohort 4 (Table S3).

ferredoxin-like peptide

str. ZAS-1

str. ZAS-2

catechol 2,3-dioxygenase

str. ZAS-1

str. ZAS-2

2-hydroxymuconic semialdehyde hydrolase

str. ZAS-1

str. ZAS-2

2-oxopent-4-enoate hydratase

str. ZAS-1

str. ZAS-2

4-hydroxy-2-oxopentanoate aldolase

str. ZAS-1

str. ZAS-2

acetaldehyde dehydrogenase

str. ZAS-1

str. ZAS-2

Fig. S1 PFAM domains and conserved functional residues and sequence motifs in *Treponema primitia* str. ZAS-1 and ZAS-2 *meta*-cleavage pathway proteins. The primary structure spans of *meta*-pathway proteins are represented by black lines, and PFAM domains within the *meta*-pathway proteins are represented by grey rectangles. Symbols representing conserved functional residues are centered over the location of the functional residues. Active site residues are represented by black inverted triangles, metal-binding residues are represented by white diamonds, residues contributing to protein structure are represented by light grey squares, and substrate-binding residues are represented by grey circles. Conserved sequence motifs are represented by white bars.

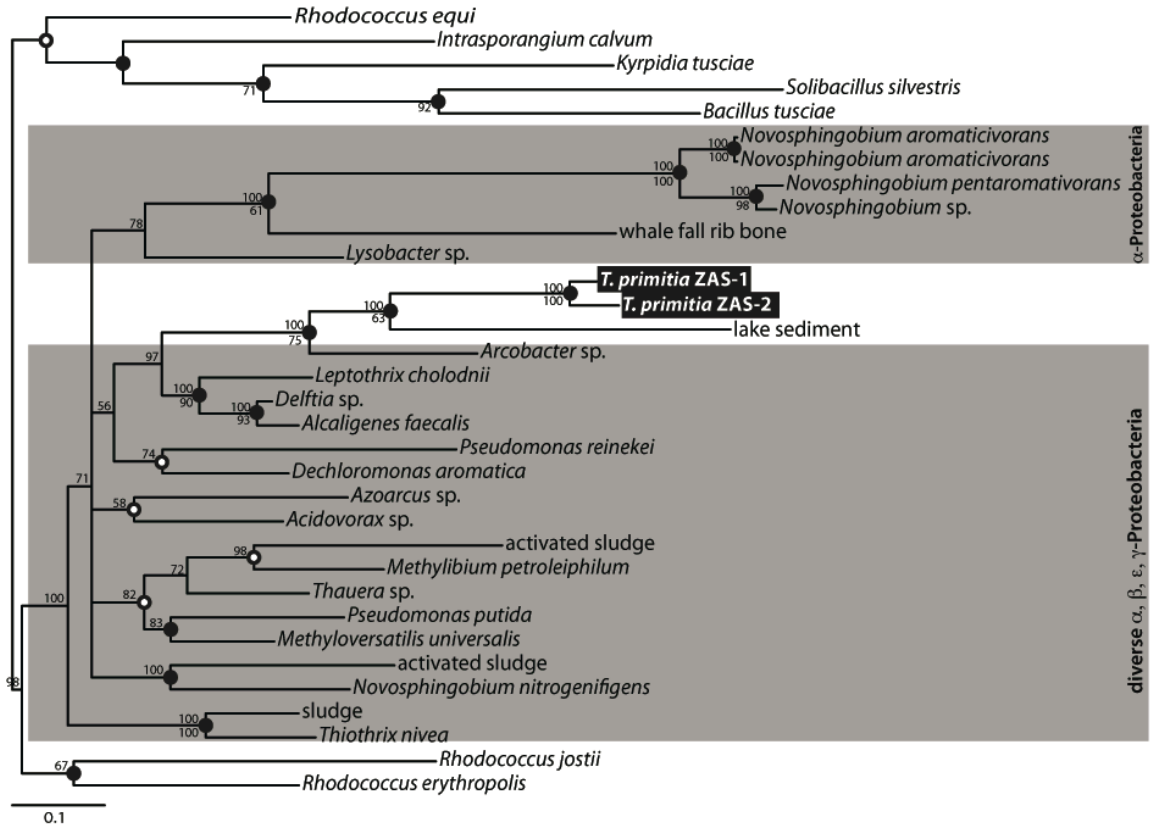

Fig. S2 Phylogenetic position of *Treponema primitia* str. ZAS-1 and ZAS-2 2-hydroxymuconic semialdehyde hydrolase (PF00561, step 2). Bayesian protein phylogenetic analysis (2540 trees from 1,016,000 generations; PSRF = 1.000; average standard deviation of split frequencies = 0.011955) is based on 222 unambiguously aligned amino acid positions of a 274 amino acid-long protein. Bayesian posterior probabilities (when greater than 50) are reported above the nodes. Phylip PROTPARS maximum parsimony support after analysis of 1000 bootstraps (when greater than 50%) is reported below nodes. Shaded circles indicate nodes supported by both maximum parsimony and Fitch distance matrix methods. Open circles indicate nodes supported by one of those methods. Scale bar indicates distance depicted as 0.1 amino acid changes per alignment position.

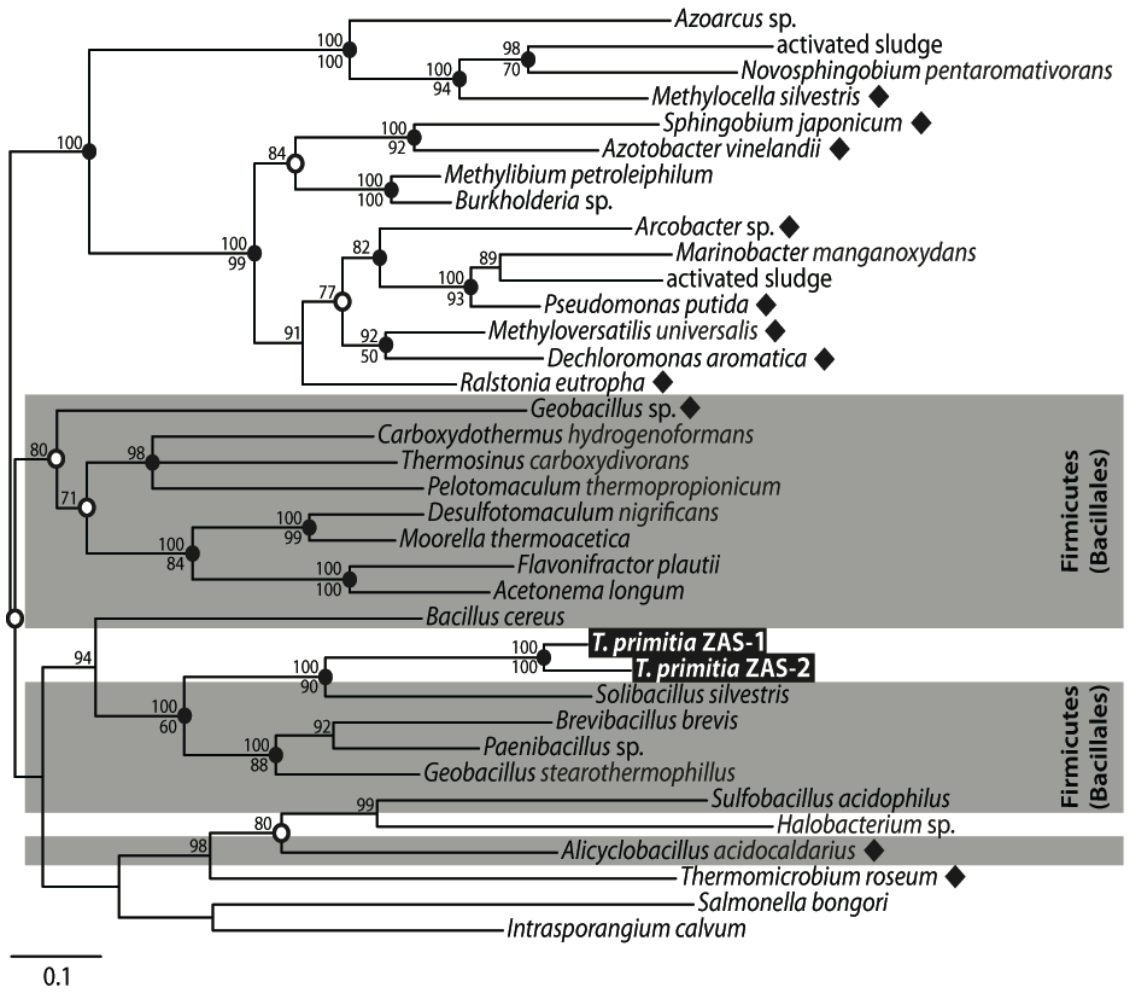

Fig. S3 Phylogenetic position of *Treponema primitia* str. ZAS-1 and ZAS-2 2-oxopent-4-enoate hydratase (PF01557, step 3). Bayesian protein phylogenetic analysis (313 trees from 125,000 generations; PSRF = 1.000; average standard deviation of split frequencies = 0.009239) is based on 243 unambiguously aligned amino acid positions of a 260 amino acid-long protein. Bayesian posterior probabilities (when greater than 50) are reported above the nodes. Phylip PROTPARS maximum parsimony support after analysis of 1000 bootstraps (when greater than 50%) is reported below nodes. Shaded circles indicate nodes supported by both maximum parsimony and Fitch distance matrix methods. Open circles indicate nodes supported by one of those methods. Scale bar indicates distance depicted as 0.1 amino acid changes per alignment position.

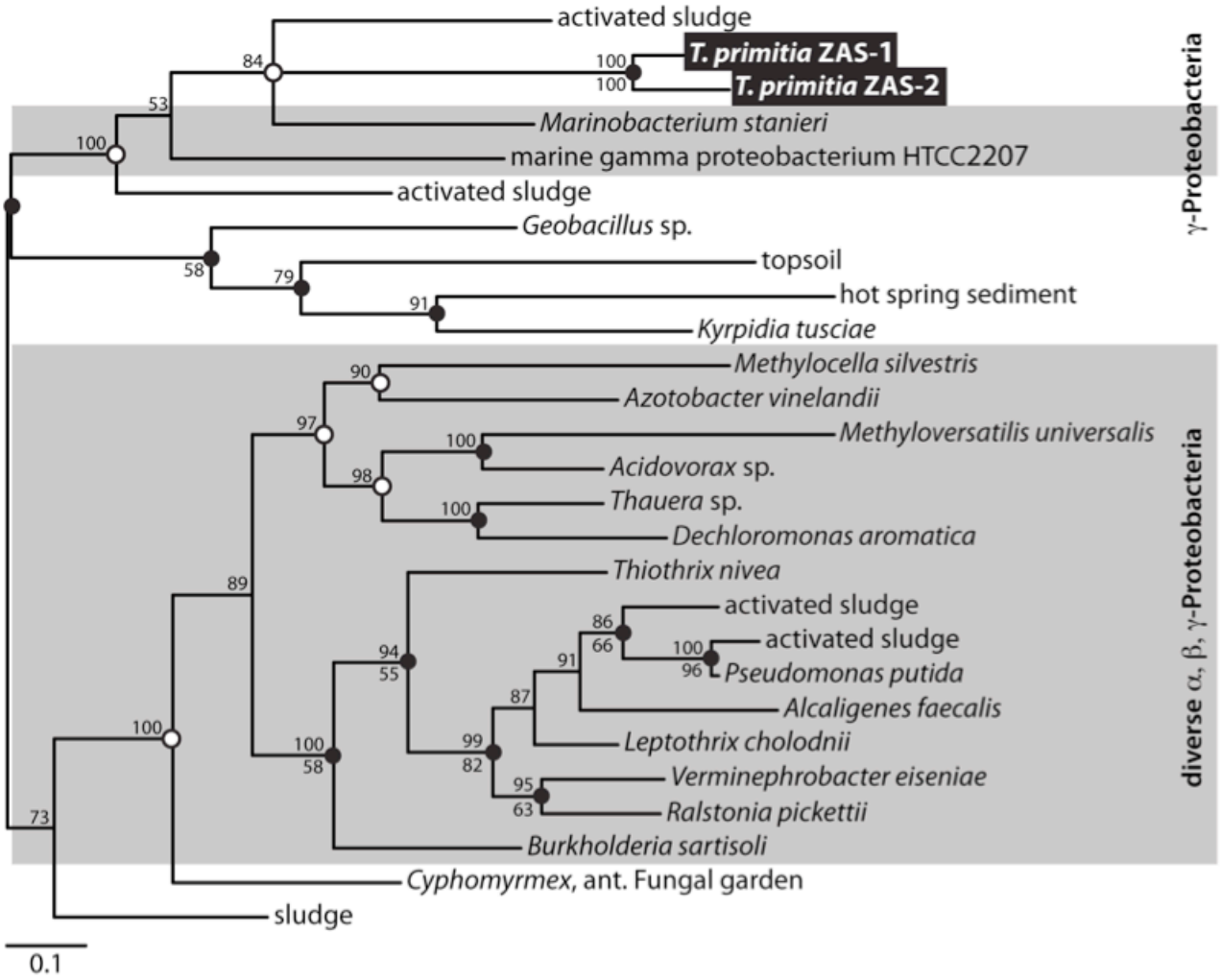

Fig. S4 Phylogenetic position of *Treponema primitia* str. ZAS-1 and ZAS-2 ferredoxin-like peptide (PF00111, Step 1'). Bayesian protein phylogenetic analysis (183 trees from 73,000 generations; PSRF = 1.001; average standard deviation of split frequencies = 0.011280) is based on 85 unambiguously aligned amino acid positions of a 98 amino acid-long protein. Bayesian posterior probabilities (when greater than 50) are reported above the nodes. Phylip PROTPARS maximum parsimony support after analysis of 1000 bootstraps (when greater than 50%) is reported below nodes. Shaded circles indicate nodes supported by both maximum parsimony and Fitch distance matrix methods. Open circles indicate nodes supported by one of those methods. Scale bar indicates distance depicted as 0.1 amino acid changes per alignment position.

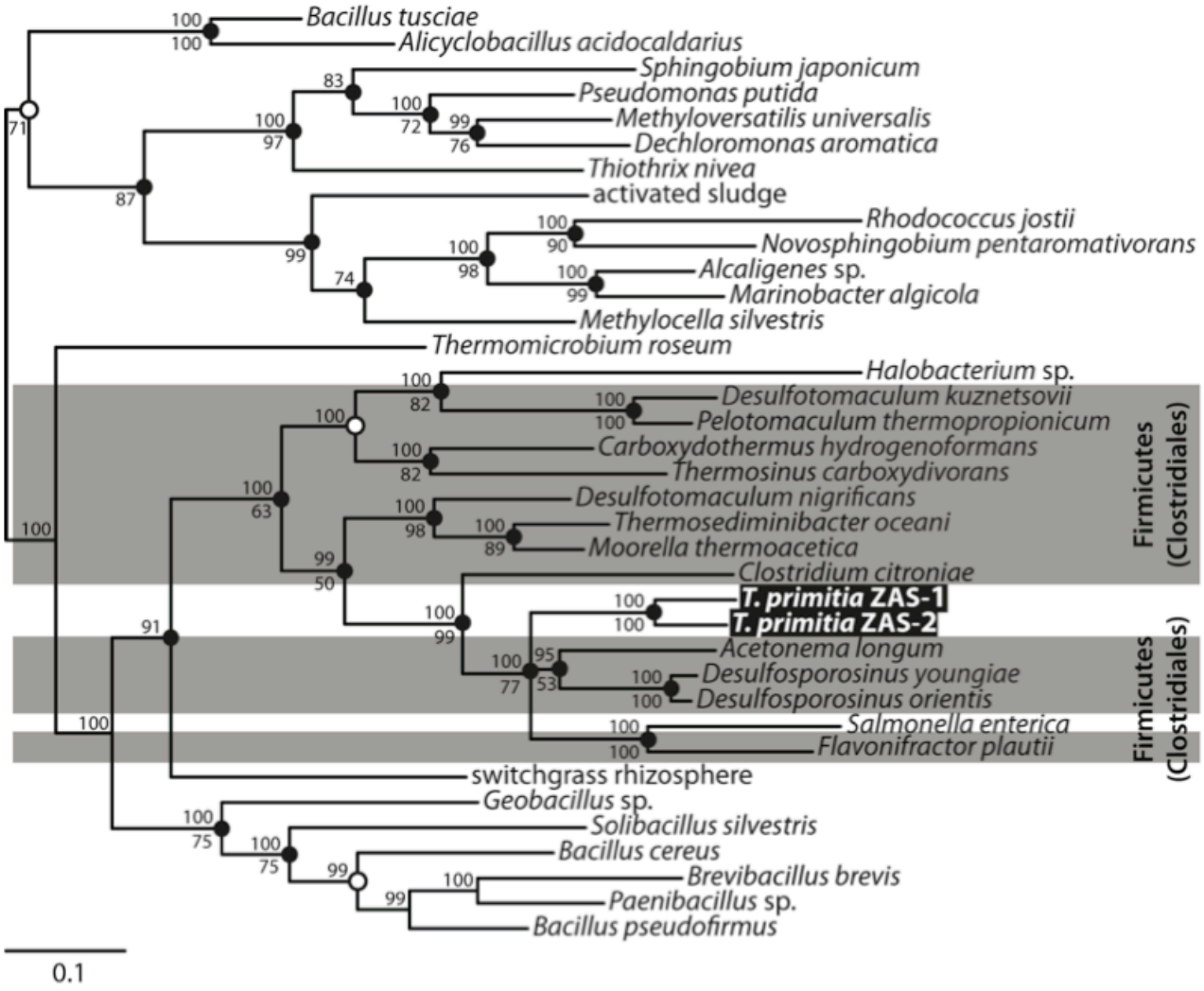

Fig. S5 Phylogenetic position of *Treponema primitia* str. ZAS-1 and ZAS-2 4-hydroxy-2-oxopentanoate aldolase (PF00682 and PF07836, Step 4). Bayesian protein phylogenetic analysis (375 trees from 150,000 generations; PSRF = 1.000; average standard deviation of split frequencies = 0.002661) is based on 331 unambiguously aligned amino acid positions of a 340 amino acid-long protein. Bayesian posterior probabilities (when greater than 50) are reported above the nodes. Phylip PROTPARS maximum parsimony support after analysis of 1000 bootstraps (when greater than 50%) is reported below nodes. Shaded circles indicate nodes supported by both maximum parsimony and Fitch distance matrix methods. Open circles indicate nodes supported by one of those methods. Scale bar indicates distance depicted as 0.1 amino acid changes per alignment position.

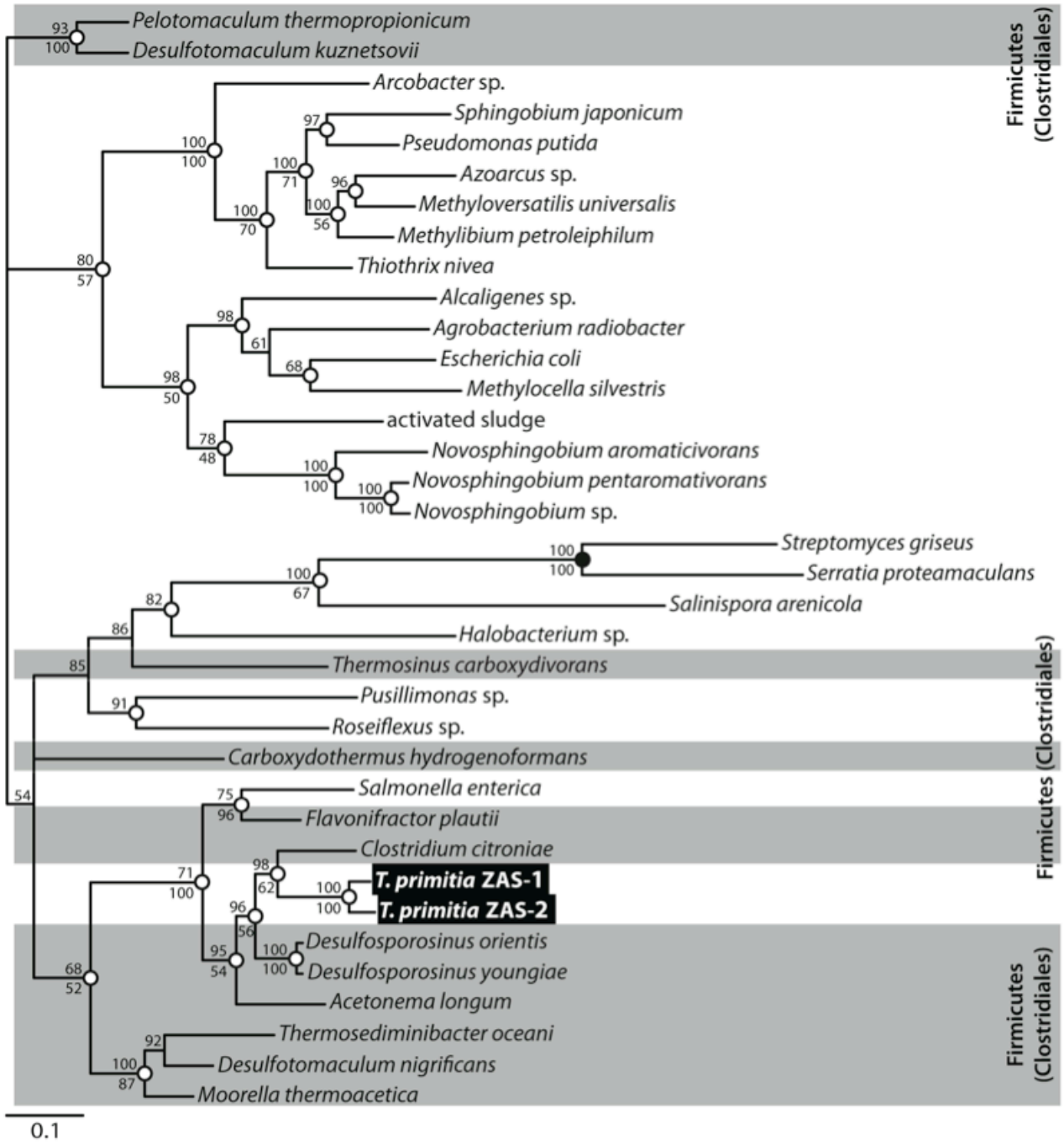

Fig. S6 Phylogenetic position of *Treponema primitia* str. ZAS-1 and ZAS-2 acetaldehyde dehydrogenase (PF01118 and PF09290, Step 5). Bayesian protein phylogenetic analysis (279 trees from 111,500 generations; PSRF = 1.004; average standard deviation of split frequencies = 0.018287) is based on 259 unambiguously aligned amino acid positions of a 291 amino acid-long protein. Bayesian posterior probabilities (when greater than 50) are reported above the nodes. Phylip PROTPARS maximum parsimony support after analysis of 1000 bootstraps (when greater than 50%) is reported below nodes. Shaded circles indicate nodes

supported by both maximum parsimony and Fitch distance matrix methods. Open circles indicate nodes supported by one of those methods. Scale bar indicates distance depicted as 0.1 amino acid changes per alignment position.

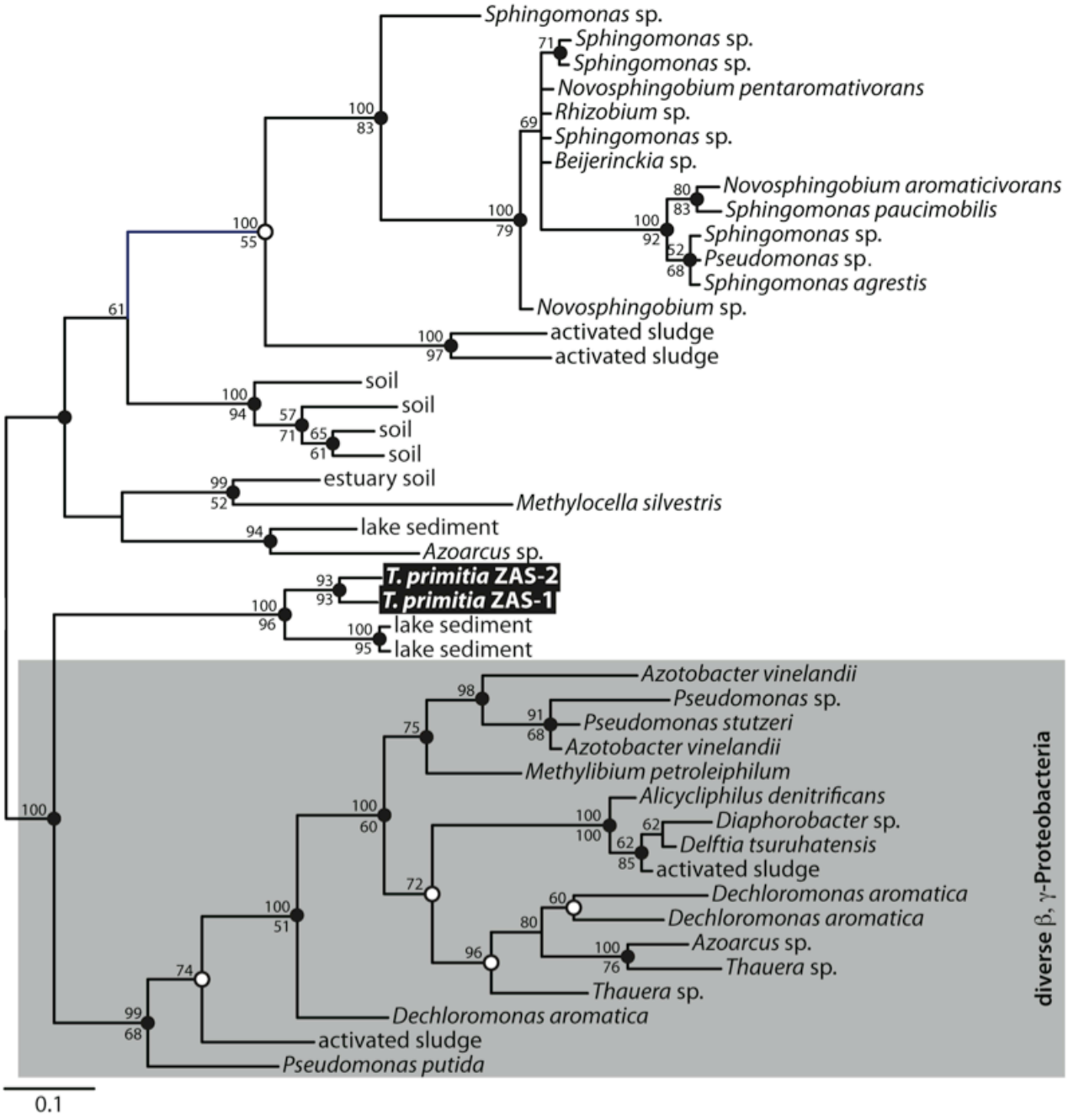

Fig. S7 Phylogenetic position of *Treponema primitia* str. ZAS-1 and ZAS-2 N-terminal Domain (PF00903). Bayesian protein phylogenetic analysis (220 trees from 88,000 generations; PSRF = 0.999; average standard deviation of split frequencies = 0.014790) is based on 56 unambiguously aligned amino acid positions of a 64 amino acid-long domain. Bayesian posterior probabilities (when greater than 50) are reported above the nodes. Phylip PROTPARS maximum parsimony support after analysis of 1000 bootstraps (when greater than 50%) is reported below nodes. Shaded circles indicate nodes supported by both maximum parsimony and Fitch distance matrix methods. Open circles indicate nodes supported by one of those

methods. Scale bar indicates distance depicted as 0.1 amino acid changes per alignment position.

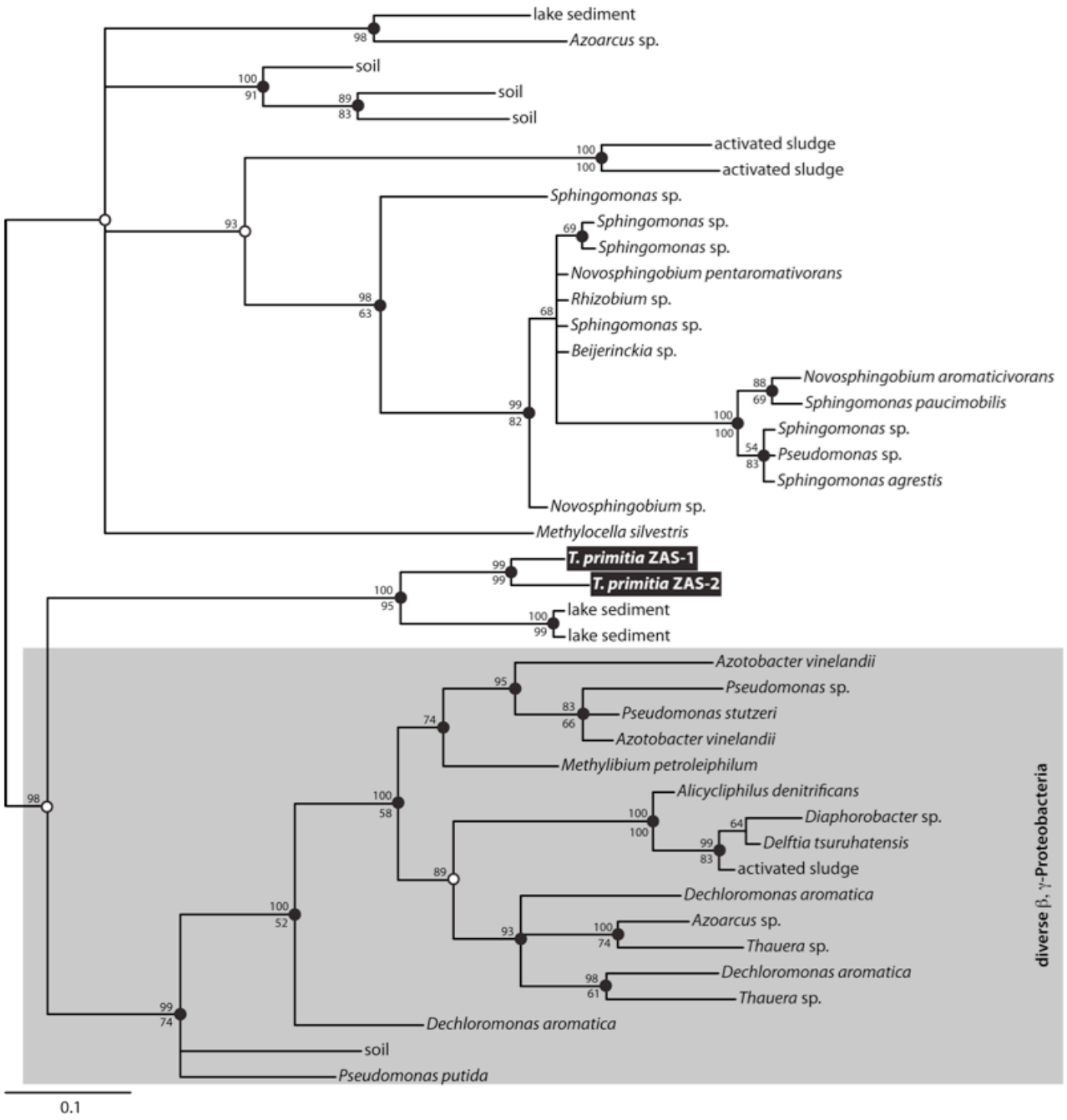

Fig. S8 Phylogenetic position of *Treponema primitia* str. ZAS-1 and ZAS-2 N-terminal Domain (PF00903) with extra-domain region. Bayesian protein phylogenetic analysis (600 trees from 240,000 generations; PSRF = 1.000; average standard deviation of split frequencies = 0.010855) is based on 67 unambiguously aligned amino acid positions of a 68 amino acid-long region. Bayesian posterior probabilities (when greater than 50) are reported above the nodes. Phylip PROTPARS maximum parsimony support after analysis of 1000 bootstraps (when greater than 50%) is reported below nodes. Shaded circles indicate nodes supported by both maximum parsimony and Fitch distance matrix methods. Open

circles indicate nodes supported by one of those methods. Scale bar indicates distance depicted as 0.1 amino acid changes per alignment position.

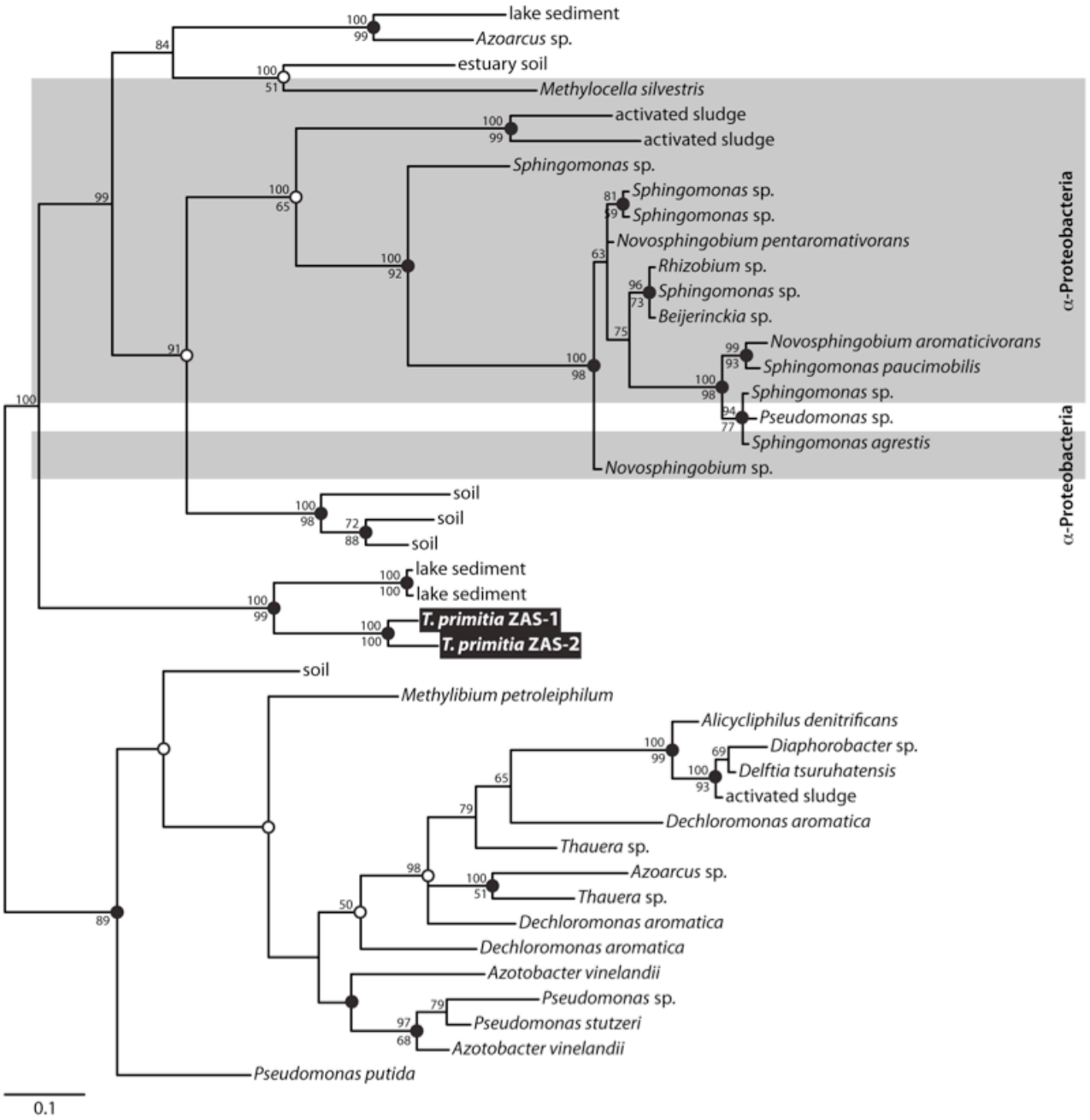

Fig. S9 Phylogenetic position of *Treponema primitia* str. ZAS-1 and ZAS-2 N-terminal Domain (PF00903) with intra-domain region. Bayesian protein phylogenetic analysis (375 trees from 150,000 generations; PSRF = 1.000; average standard deviation of split frequencies = 0.010443) is based on 88 unambiguously aligned amino acid positions of a 139 amino acid-long region. Bayesian posterior probabilities (when greater than 50) are reported above the nodes. Phylip PROTPARS maximum parsimony support after analysis of 1000 bootstraps (when greater than 50%) is reported below nodes. Shaded circles indicate nodes supported by both maximum parsimony and Fitch distance matrix methods. Open

circles indicate nodes supported by one of those methods. Scale bar indicates distance depicted as 0.1 amino acid changes per alignment position.

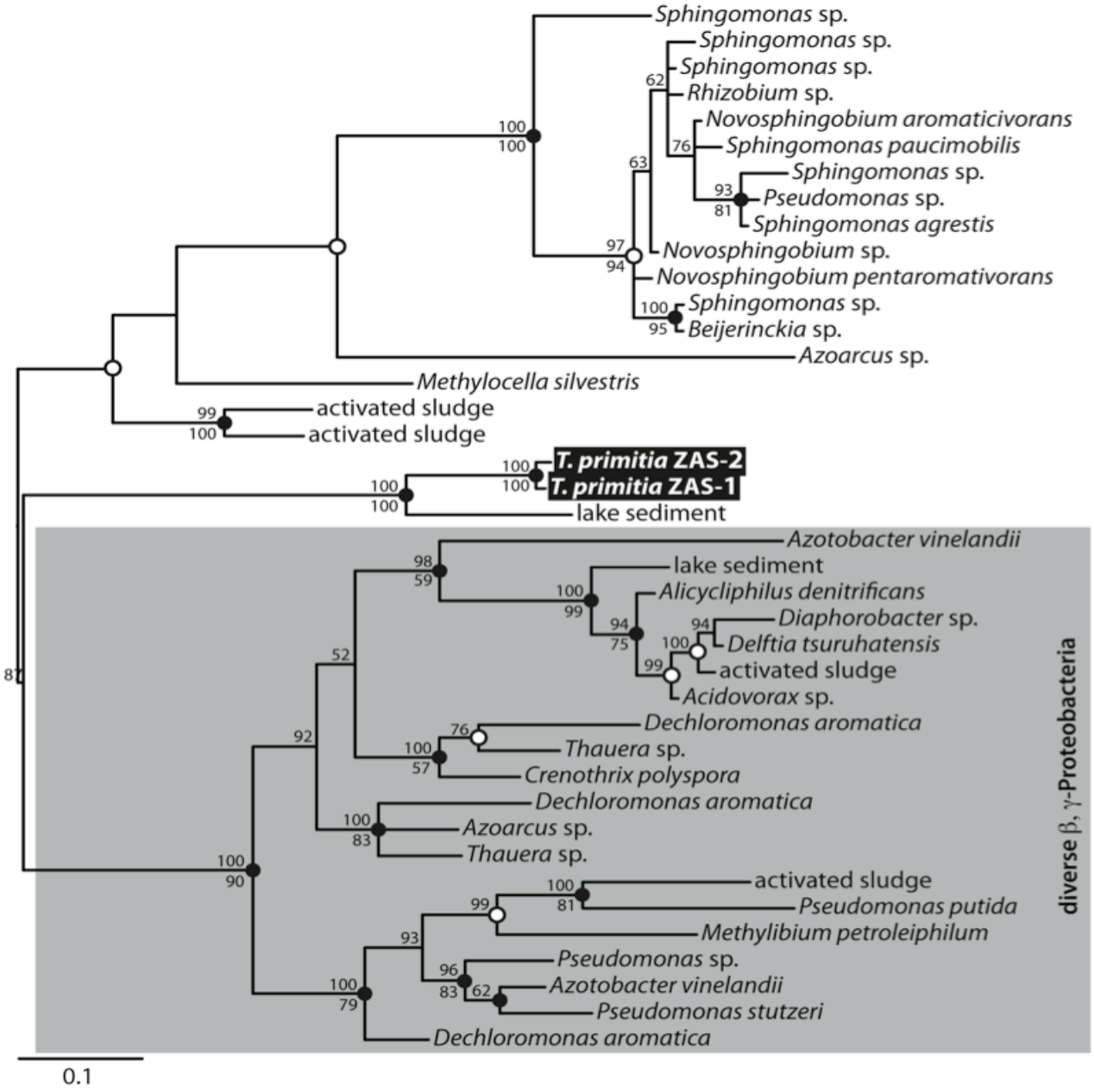

Fig. S10 Phylogenetic position of *Treponema primitia* str. ZAS-1 and ZAS-2 C-terminal Domain (PF00903). Bayesian protein phylogenetic analysis (350 trees from 140,000 generations; PSRF = 1.000; average standard deviation of split frequencies = 0.014588) is based on 115 unambiguously aligned amino acid positions of a 116 amino acid-long domain. Bayesian posterior probabilities (when greater than 50) are reported above the nodes. Phylip PROTPARS maximum parsimony support after analysis of 1000 bootstraps (when greater than 50%) is reported below nodes. Shaded circles indicate nodes supported by both maximum parsimony and Fitch distance matrix methods. Open circles indicate nodes supported by one of those methods. Scale bar indicates distance depicted as 0.1 amino acid changes per alignment position.

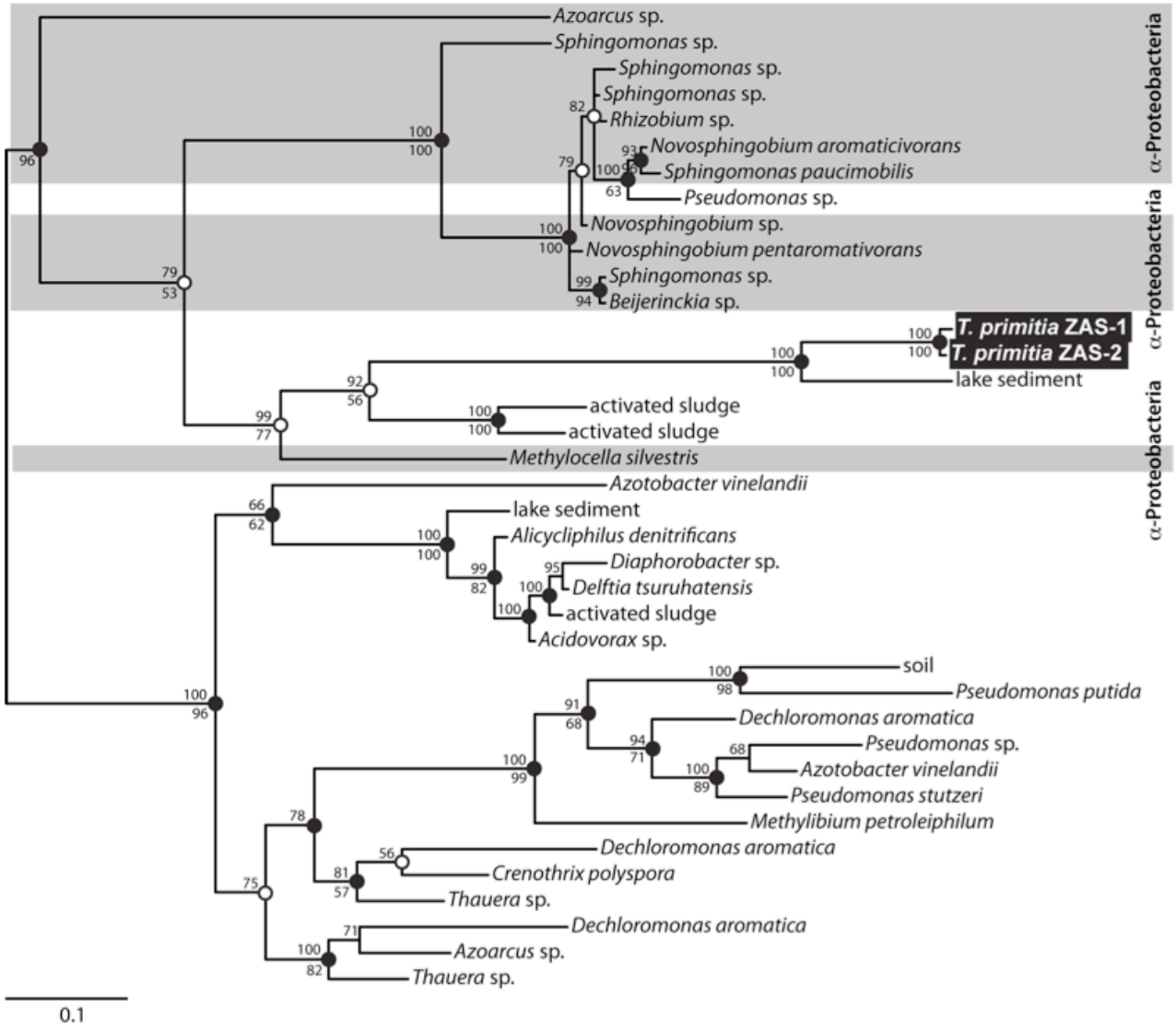

Fig. S11 Phylogenetic position of *Treponema primitia* str. ZAS-1 and ZAS-2 C-terminal Domain (PF00903) with extra-domain region. Bayesian protein phylogenetic analysis (450 trees from 180,000 generations; PSRF = 1.000; average standard deviation of split frequencies = 0.010860) is based on 155 unambiguously aligned amino acid positions of a 158 amino acid-long region. Bayesian posterior probabilities (when greater than 50) are reported above the nodes. Phylip PROTPARS maximum parsimony support after analysis of 1000 bootstraps (when greater than 50%) is reported below nodes. Shaded circles indicate nodes supported by both maximum parsimony and Fitch distance matrix methods. Open circles indicate nodes supported by one of those methods. Scale bar indicates distance depicted as 0.1 amino acid changes per alignment position.

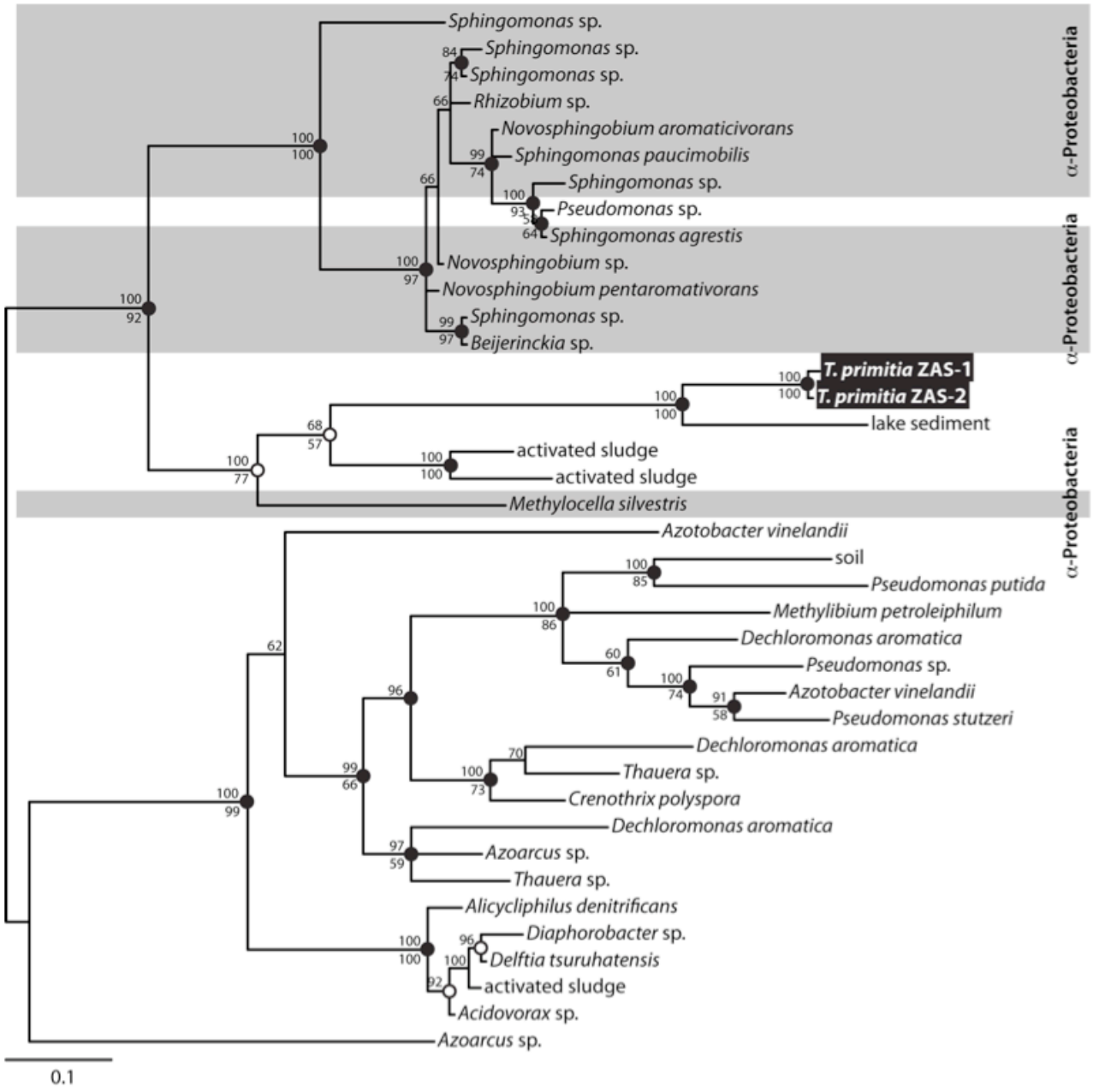

Fig. S12 Phylogenetic position of *Treponema primitia* str. ZAS-1 and ZAS-2 C-terminal Domain (PF00903) with intra-domain region. Bayesian protein phylogenetic analysis (500 trees from 200,000 generations; PSRF = 1.000; average standard deviation of split frequencies = 0.010844) is based on 140 unambiguously aligned amino acid positions of a 141 amino acid-long region. Bayesian posterior probabilities (when greater than 50) are reported above the nodes. Phylip PROTPARS maximum parsimony support after analysis of 1000 bootstraps (when greater than 50%) is reported below nodes. Shaded circles indicate nodes supported by both maximum parsimony and Fitch distance matrix methods. Open circles indicate nodes supported by one of those methods. Scale bar indicates distance depicted as 0.1 amino acid changes per alignment position.

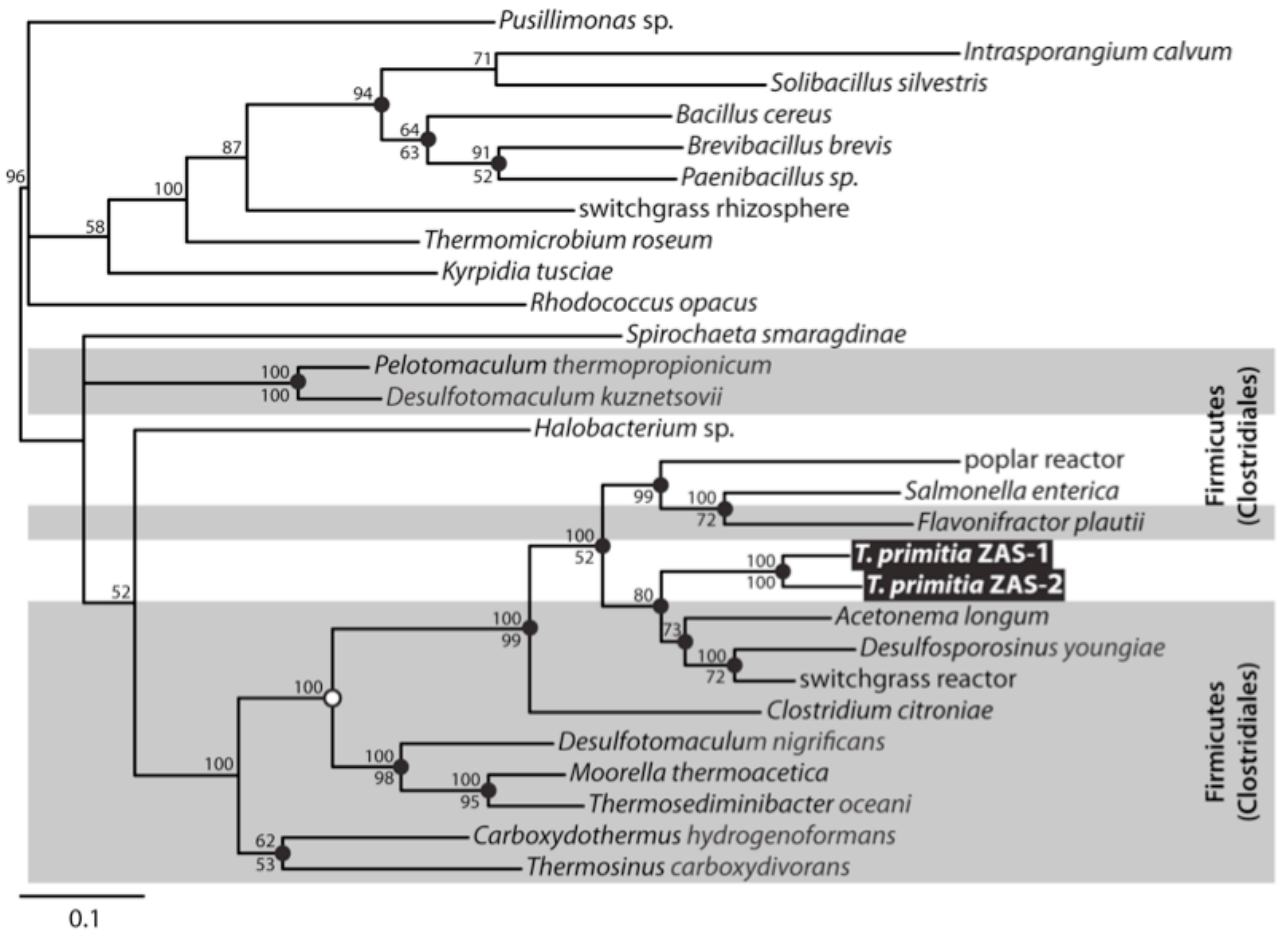

Fig. S13 Phylogenetic position of *Treponema primitia* str. ZAS-1 and ZAS-2 4-hydroxy-2-oxopentanoate aldolase HMGL-like domain (PF00682). Bayesian protein phylogenetic analysis (313 trees from 125,000 generations; PSRF = 1.000; average standard deviation of split frequencies = 0.009037) is based on 243 unambiguously aligned amino acid positions of a 232 amino acid-long domain. Bayesian posterior probabilities (when greater than 50) are reported above the nodes. Phylip PROTPARS maximum parsimony support after analysis of 1000 bootstraps (when greater than 50%) is reported below nodes. Shaded circles indicate nodes supported by both maximum parsimony and Fitch distance matrix methods. Open circles indicate nodes supported by one of those methods. Scale bar indicates distance depicted as 0.1 amino acid changes per alignment position.

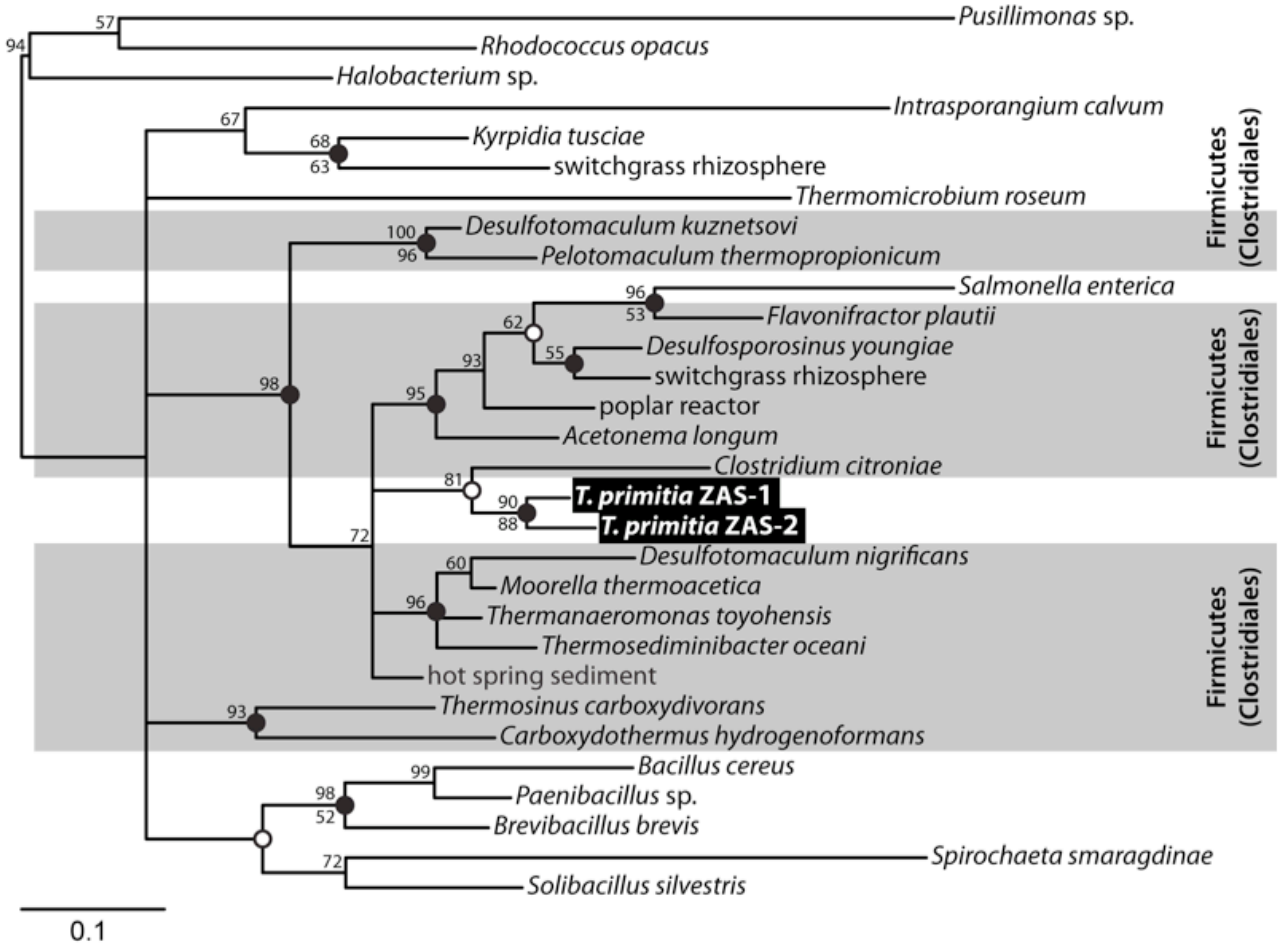

Fig. S14 Phylogenetic position of *Treponema primitia* str. ZAS-1 and ZAS-2 4-hydroxy-2-oxopentanoate aldolase DmpG-like communication domain (PF07836). Bayesian protein phylogenetic analysis (480 trees from 192,000 generations; PSRF = 1.000; average standard deviation of split frequencies = 0.016899) is based on 65 unambiguously aligned amino acid positions of a 66 amino acid-long domain. Bayesian posterior probabilities (when greater than 50) are reported above the nodes. Phylip PROTPARS maximum parsimony support after analysis of 1000 bootstraps (when greater than 50%) is reported below nodes. Shaded circles indicate nodes supported by both maximum parsimony and Fitch distance matrix methods. Open circles indicate nodes supported by one of those methods. Scale bar indicates distance depicted as 0.1 amino acid changes per alignment position.

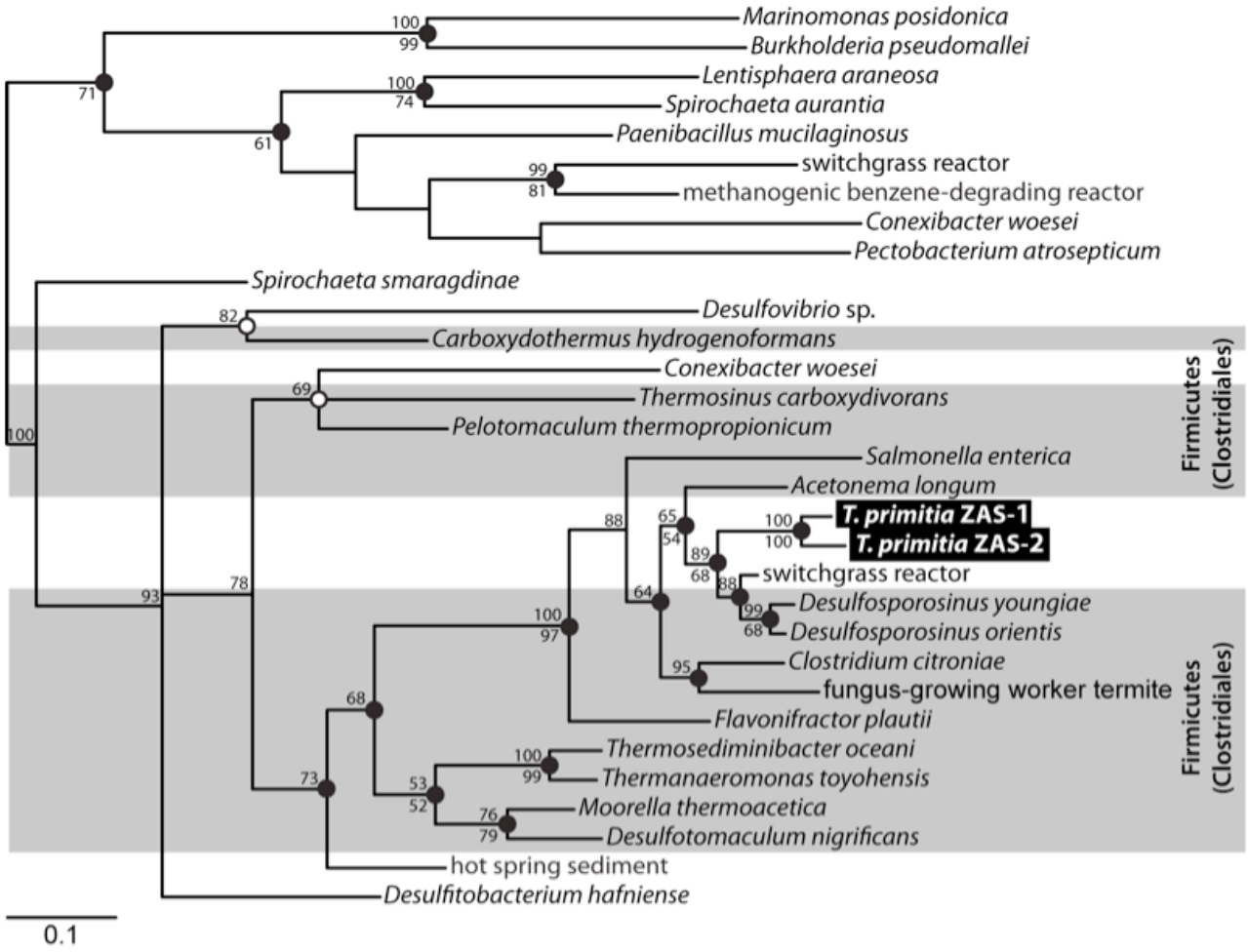

Fig. S15 Phylogenetic position of *Treponema primitia* str. ZAS-1 and ZAS-2 acetaldehyde dehydrogenase, NAD binding domain (PF01118). Bayesian protein phylogenetic analysis (123 trees from 205,000 generations; PSRF = 1.000; average standard deviation of split frequencies = 0.010781) is based on 110 unambiguously aligned amino acid positions of a 110 amino acid-long domain. Bayesian posterior probabilities (when greater than 50) are reported above the nodes. Phylip PROTPARS maximum parsimony support after analysis of 1000 bootstraps (when greater than 50%) is reported below nodes. Shaded circles indicate nodes supported by both maximum parsimony and Fitch distance matrix methods. Open circles indicate nodes supported by one of those methods. Scale bar indicates distance depicted as 0.1 amino acid changes per alignment position.

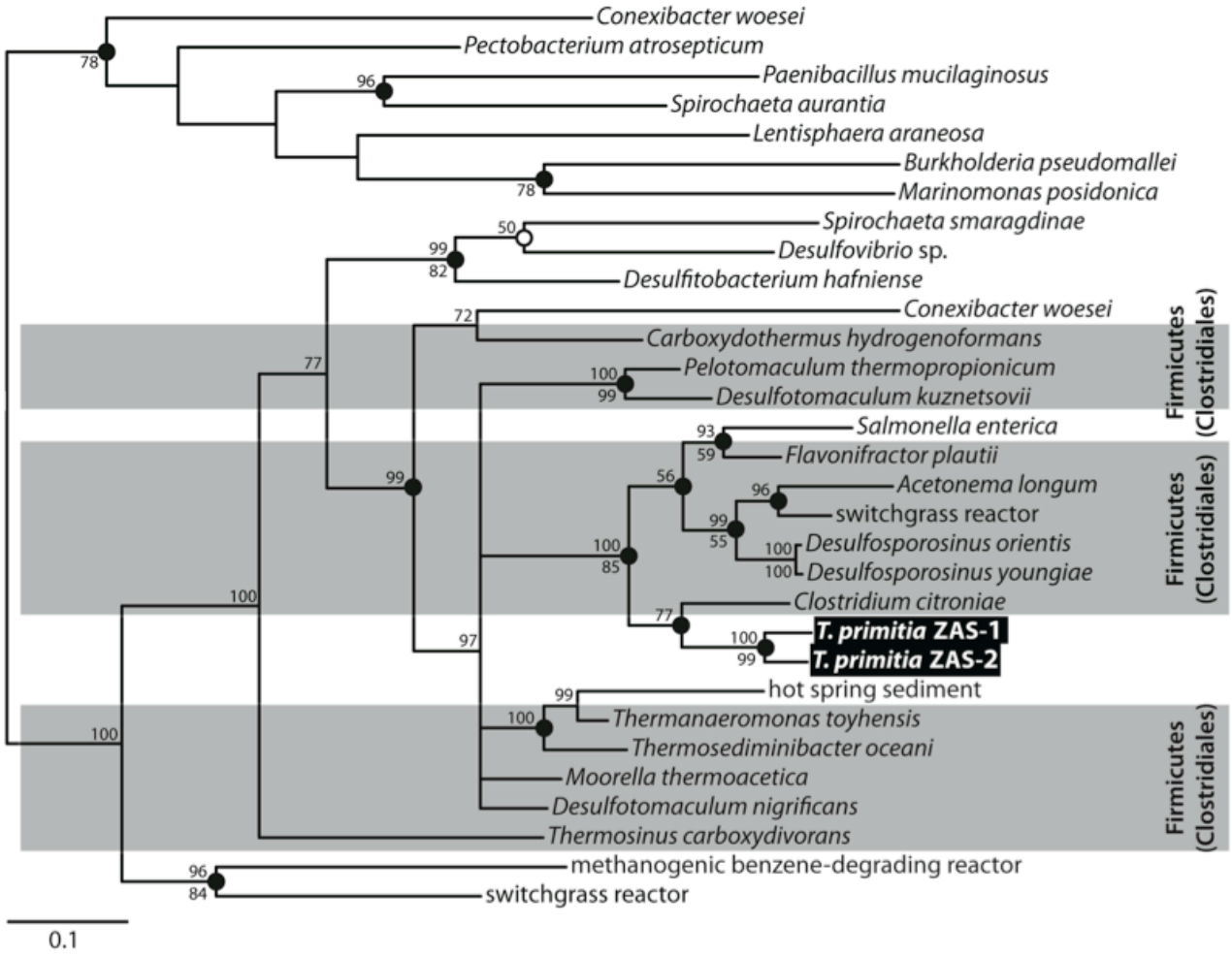

Fig. S16 Phylogenetic position of *Treponema primitia* str. ZAS-1 and ZAS-2 acetaldehyde dehydrogenase dimerisation domain (PF09290). Bayesian protein phylogenetic analysis (145 trees from 58,000 generations; PSRF = 0.999; average standard deviation of split frequencies = 0.013476) is based on 135 unambiguously aligned amino acid positions of a 137 amino acid-long domain. Bayesian posterior probabilities (when greater than 50) are reported above the nodes. Phylip PROTPARS maximum parsimony support after analysis of 1000 bootstraps (when greater than 50%) is reported below nodes. Shaded circles indicate nodes supported by both maximum parsimony and Fitch distance matrix methods. Open circles indicate nodes supported by one of those methods. Scale bar indicates distance depicted as 0.1 amino acid changes per alignment position.

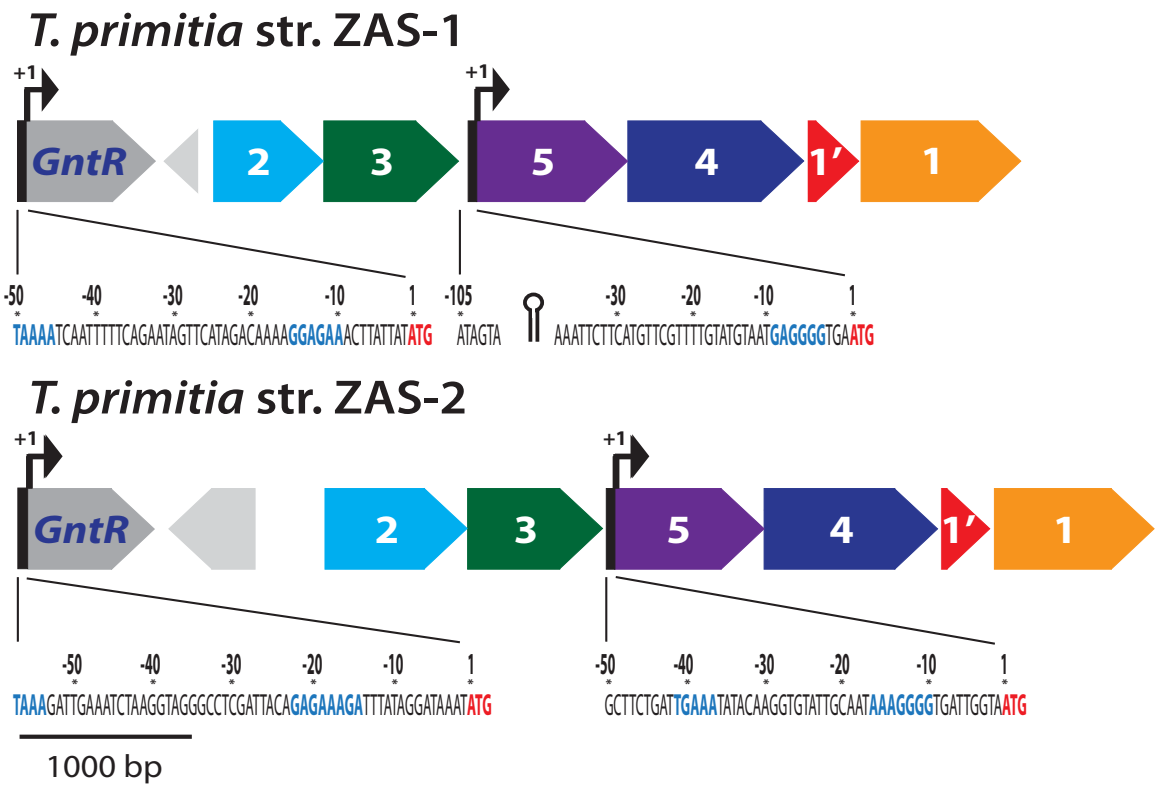

Fig. S17 Transcriptional regulatory elements within the *Treponema primitia* str. ZAS-1 and str. ZAS-2 *meta*-cleavage pathway gene neighborhoods. Promoter regions are black rectangles and arrows indicate direction of transcription. Nucleotide bases of the promoter region are shown with conserved promoter elements in blue and transcriptional start sites in red. Stem loop structure is represented by a hairpin-like symbol. Ferredoxin-like peptide (step 1', red), catechol 2,3-dioxygenase (step 1, orange), 2-hydroxymuconic semialdehyde hydrolase (step 2, light blue), 2-oxopent-4-enoate hydratase (step 3, dark green), 4-hydroxy-2-oxopentanoate aldolase (step 4, dark blue), acetaldehyde dehydrogenase (step 5, purple), putative regulatory genes (dark grey), and hypothetical proteins (light grey) are depicted. Regulatory families are noted.

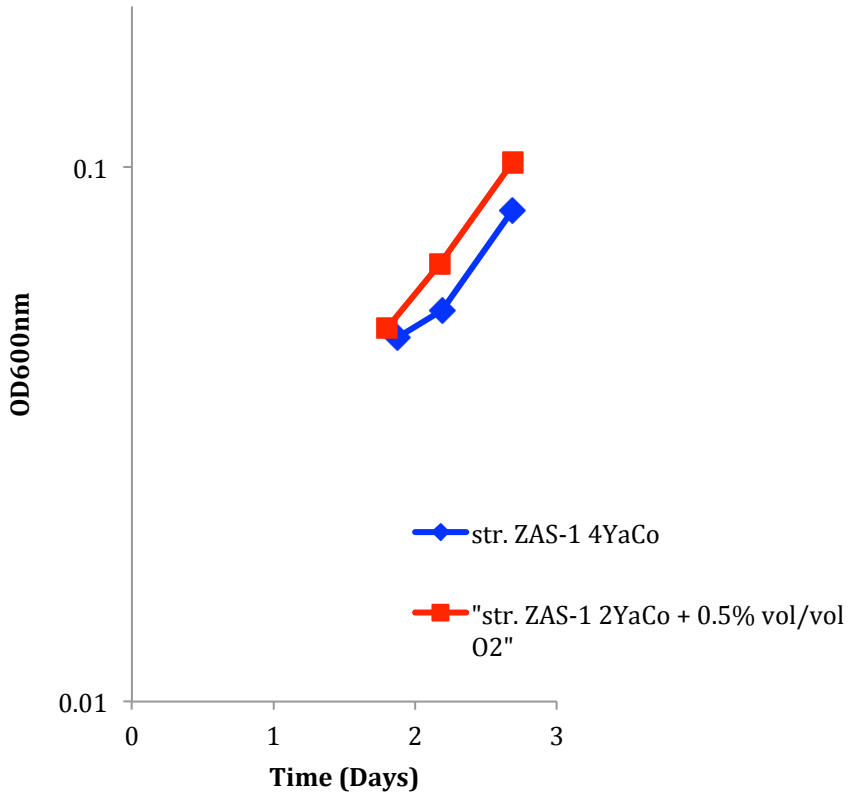

Fig. S18: Microoxic growth of *Treponema primitia* str. ZAS-1. *T. primitia* str. ZAS-1 grew in 5mL of 2YACo, DTT-free media, with an 80% N_2/20% CO_2 headspace to which air was added sterilely to achieve 0.5% vol/vol O_2 (red) and 5mL of 4YACo, DTT-reduced media with an 80% H_2/20% CO_2 headspace (blue). Both treatment types were supplemented with a vitamin mixture (12-vitamin, vitamin B12, and folinic acid solutions, Graber & Breznak, 2004). Cultures were contained in 25mL butyl rubber-stoppered Balch tunes and incubated at 25°C horizontally and still. First three days of growth are shown.